GENERAL DETERRENT EFFECTS OF POLICE PATROL IN CRIME "HOT SPOTS": A RANDOMIZED, CONTROLLED TRIAL*

LAWRENCE W. SHERMAN
University of Maryland

DAVID WEISBURD
Hebrew University

Many criminologists doubt that the dosage of uniformed police patrol causes any measurable difference in crime. This article reports a one-year randomized trial in Minneapolis of increases in patrol dosage at 55 of 110 crime "hot spots," monitored by 7,542 hours of systematic observations. The experimental group received, on average, twice as much observed patrol presence, although the ratio displayed wide seasonal fluctuation. Reductions in total crime calls ranged from 6 percent to 13 percent. Observed disorder was only half as prevalent in experimental as in control hot spots. We conclude that substantial increases in police patrol presence can indeed cause modest reductions in crime and more impressive reductions in disorder within high crime locations.

In 1974 the Kansas City Preventive Patrol Experiment (Kelling et al. 1974a) shook the theoretical foundations of American policing. The year-long study found that experimentally manipulated variations in the dosage of police patrol across 15 patrol beats had virtually no statistically significant effects on street crime. Then-Kansas

City Police Chief Joseph McNamara concluded that "routine preventive patrol in marked police cars has little value in preventing crime or making citizens feel safe".

This finding has dominated police thinking about patrol strategies for more than two decades. Despite contradictory evidence from studies employing equally rigorous research designs (Chaiken 1978; Press 1971; Schnelle et al. 1977; Sherman 1990), the Kansas City finding remains the most influential test of the general deterrent effects of patrol on crime. It has convinced many distinguished scholars that no matter how it is deployed, police presence does not deter. Klockars (1983:130), for example, concludes that "it makes about as much sense to have police patrol routinely in cars to fight crime as it does to have firemen patrol routinely in fire trucks to fight fire." Skolnick and Bayley (1986:4) conclude that "random motor patrolling neither reduces crime nor improves chances of catching suspects." Gottfredson and Hirschi (1990:270) conclude that "no evidence exists that augmentation of police forces or equipment, differential patrol strategies, or differential intensities of surveillance have an effect on crime rates." Even Felson (1994:10-11), a rational choice theorist, interprets the Kansas City findings as evidence that "patrol has no impact on crime rates" because the low density of modern metropolitan areas makes police presence a "drop in the bucket."

The Kansas City experiment does not justify such strong conclusions. Years of debate have revealed substantial statistical, measurement, and conceptual problems in its design. The *statistical* problem is the bias, found in most area-level designs, toward the null hypothesis; the weak statistical power of such designs makes it very difficult to find an effect of patrol (or any other intervention) even when such an effect may be present (Fienberg, et al 1976). The *measurement* problem lies in determining exactly how much dosage was delivered in each of the experimental conditions, which the Kansas City study did not do. Both of these issues point up Felson's conceptual problem of dosage levels: the premise that large patrol beats or neighborhoods are the appropriate unit for allocating and testing the impact of patrol, which dilutes available dosage too much to make a reasonable impact likely (Farrington 1982).

In this article we explore those problems and a research solution: the use of very small clusters of high-crime addresses ("hot spots") as the unit of analysis instead of patrol beats or neighborhoods. We then present the research design and the results of a test of the general deterrent effects of patrol in hot spots.[1]

---

[1] The use of "general" refers to potential offenders, in contrast to "specific" deterrence of future crime by persons who have been punished in the past. We do

RESEARCH DESIGN ISSUES IN PATROL AND CRIME

Statistical Bias towards the Null Hypothesis

The major statistical limitation in all experiments in patrol beat or neighborhood-level crime reduction is lack of power (Freiman et al 1978; Sherman 1986:362-64; Zimring 1978:162-63). This problem has three dimensions, each of which creates a bias against demonstrating any impact of policing (or other interventions) on crime. One statistical power issue is the low frequency of *crimes* in most neighborhoods. A second is the number of *citizens* who must be interviewed in each community to permit reliable estimates of changes in the victimization rate of that community. The third is the number of *communities* included in community-level tests of policing strategies.

Most patrol beat-sized neighborhoods in most cities suffer relatively few serious crimes each year. To provide a reliable estimate of the prevalence of most types of crime through victimization surveys, large samples must be drawn for each area. The expense entailed in drawing these samples is so great that it limits the number of areas which can be studied at reasonable cost. Measures of reported crime are less expensive to collect, but they also provide low base rates. One robbery (or less) per month, for example, is a common rate for many patrol beats, as it was in the San Diego Field Interrogation Experiment in beats of 7,000 to 14,000 residents (Boydstun 1975:16, 32). That rarity creates a bias toward the null hypothesis for any crime-specific statistical tests of the impact of interventions. Kelling et al (1974b:96), for example, found that a 300 percent increase in reported robberies in the less heavily patrolled areas was not statistically significant because the large relative difference reflected an absolute difference of less than one outside robbery per month. The observed difference in robbery in Kansas City might have been significant with a sample size of hundreds of patrol beats. Few cities of over 250,000, however, have even 50 patrol beats, let alone hundreds.

Measuring and Varying Patrol Dosage Levels

A substantive bias toward the null hypothesis in the Kansas City design may have been created by insufficient differences in patrol dosage. Larson (1975) argued that five factors created as much visible patrol presence in the unpatrolled beats as would normal patrol dosage (but see Pate and Kelling 1975): 1) travel into and out

---

not imply that general deterrence of crime in hot spots necessarily deters crime "generally" throughout the city beyond the hot spot location; we treat that issue as empirical rather than conceptual.

of the beats to answer calls for service, 2) the operation of other (nonpatrol) units in marked cars, 3) greater use of sirens and lights, 4) more frequent responses by two units, and 5) more police-initiated contacts. This does not necessarily discount the failure of the areas with increased dosage to show more crime reduction than those with normal dosage (Zimring 1978:143). Yet it raises a key question: How certain can we be of the exact dosage of visible police presence delivered in any of the 15 beats?

If we assume that the dosage levels in Kansas City actually may have varied very little, that point alone may explain why the Kansas City results differ from those of most other quasi-experimental patrol deterrence studies. In the 1966 study of New York City police, a reported 40 percent increase in patrol car presence reduced target crimes (Press 1971). In the New York City subway study, an increase of almost 300 percent in police staffing apparently caused an initial deterrent effect (Chaiken 1978). In Nashville, a 400 percent increase in police-recorded patrol time in four target areas was associated with significant reductions in total crime (Schnelle et al. 1977). Large increases in dosage thus may be essential if any effect on crime is to be observed. The Kansas City design called for substantial increases, but could not measure the dosage reliably. In the absence of carefully measured levels of patrol dosage, it is almost impossible to interpret the Kansas City preventive patrol experiment.

The measurement and the control of dosage are closely related. Where dosage levels cannot be measured, it is difficult to advise police supervisors on whether proper levels are being delivered. It is also impossible to develop a precise dosage-response curve from multiple experiments, an essential condition for building theory. Thus the basic issues in measuring police patrol dosage must be carefully considered.

Patrol dosage can be measured from the perspective of either the police or the criminal. The police perspective on their own whereabouts can be measured through police logs or notes of independent observers riding in patrol cars. The potential criminal's perspective on police whereabouts can be measured by independent observers stationed in public places. To estimate with any precision the odds that police will pass any particular location, one would require repeated observations from a large sample of all possible observation posts within patrol car beats. The need to sample both space and time could make the gathering of such estimates even more costly per unit of analysis than personal victimization surveys—as long as the unit of analysis remained the entire low-

density patrol beat rather than the small parts of each beat where crime is concentrated.

Moreover, spreading observations over entire patrol beats would dilute the power of the observation sample to produce a reliable estimate of police presence in any given place—just as spreading patrol itself dilutes the potential deterrent threat of police presence in any one place. This point raises the more general question of the appropriate unit of analysis for patrol experiments and operations, which should guide the methods of measurement.

The Unit of Analysis: Patrol Beats or Hot Spots?

The premise of organizing patrol by beats is that crime could happen anywhere and that the entire beat must be patrolled. Computer-age data, however, have given new support to Henry Fielding's ([1751] 1977) eighteenth century proposal that police pay special attention to a small number of locations at high risk of crime. If only 3 percent of the addresses in a city produce more than half of all the requests for police response, if no police cars are dispatched to 40 percent of the addresses and intersections in a city over one year, and, if among the 60 percent with any requests, the majority (31%) register only one request per year (Sherman, Gartin, and Buerger, 1989), then concentrating police in a few locations makes more sense than spreading them evenly throughout a beat (Sherman and Weisburd 1995).

The main argument against directing extra resources to the hot spots is that it would simply displace crime problems from one address to another without achieving any overall or lasting reduction in crime. The premise of this argument is that a fixed supply of criminals is seeking outlets for the fixed number of crimes they are predestined to commit. Although that argument may fit some public drug markets (Sherman 1990; but see Green 1995; Weisburd and Green 1995), it does not fit all crime or even all vice. One carefully studied prostitution market was closed by a police crackdown (and road closing) with no apparent displacement (Matthews 1986). There is no evidence that displacement is certain across all crime categories (Cornish and Clarke 1987); the most thorough study of displacement from increased patrol (Press 1971) found that the estimate of displaced crime was less than the reduction of crime in the experimental precinct (see also Barr and Pease 1990).

In any case, displacement is merely a rival theory explaining *why* crime declines at a specific hot spot, *if* it declines. The first step is to see whether crime can be reduced at those spots at all, with a research design capable of giving a fair answer to that question.

The geographic concentration of many crimes and many calls to police about crime provides a solution to all three dimensions of the statistical power problem discussed above. First, each "hot spot" cluster of visually connected addresses offers ample numbers of calls and crimes for statistical analysis of changes at that location. Second, any city contains far more hot spots than patrol beats, so there is no difficulty in constructing a large sample of hot spot locations. Third, concentrating patrol dosage in a hot spot could create a substantial increase in patrol dosage in a very small world, and would make systematic observation an economically viable way of measuring patrol dosage levels. Although this solution does not make victimization interviews more economical, it makes feasible an even more direct measure of the most frequent kinds of crime: systematic observation, which also can measure patrol presence. The design presented below demonstrates how this solution can be operationalized, and shows the resulting statistical power.

EXPERIMENTAL DESIGN

Selection of City

We designed the experiment in collaboration with the Minneapolis Police Department, where the pattern of hot spots across all offenses had first been demonstrated (Sherman et al 1989). The experiment was endorsed by a vote of the City Council upon the Mayor's recommendation, despite the predicted effect of minimizing patrols in outlying Council members' areas and concentrating police presence in the inner core of the city, where hot spots of crime were more prevalent. The experiment also required the cooperation of the entire patrol force; this was facilitated by a recent change in case law, which gave the Chief of Police more control over the four patrol precinct commanders. Police cooperation was also pursued through briefings, pizza parties, and t-shirts bearing the project's logo ("Minneapolis Hot Spot Cop").

Selection of Hot Spots

We defined hot spots operationally as small clusters of addresses with frequent "hard" crime calls as well as substantial "soft" crime calls for service (Reiss 1985).[2] We then limited the boundaries of each spot conceptually as easily visible from an epicenter (Sherman et al 1989). This definition failed to solve the problem of crimes occurring at rear entrances to addresses listed in

---

2 Examples of "hard crime" calls are holdup alarms, burglary, shooting, stabbing, auto theft, theft from autos, assault, and rape. Examples of "soft crime" calls are audible break-in alarms, disturbances, drunks, noise, unwanted persons at businesses, vandalism, prowlers, fights, and person down.

the dispatch data, but the "noise" from this problem should not threaten an internally valid comparison between two randomized groups of hot spots, both of which suffer that noise problem to roughly the same extent.

The selection procedure began with a data file on all dispatched calls for police service citywide for the most recent year before the beginning of the selection analysis (June 6, 1987 through June 5, 1988; this is described below as the "selection year," as distinct from the "baseline" year preceding the starting date of the experiment). In the selection year we identified 5,538 addresses and intersections with more than three calls to police about incidents that we defined as "hard crime". We then employed a computer mapping program, MAPINFO, to locate most of the addresses, so that inspection of the computer printouts for each map grid could reveal what appeared to be visually connected clusters of these addresses.[3] Using this technique, we identified and mapped 420 address clusters with 20 or more hard crime calls (see Buerger, Cohn, and Petrosino 1995).

All 420 of these clusters were visually inspected by field staff members. The inspections had three principal goals. One goal was to reconfigure the boundaries suggested by the computer map to make them consistent with the definition based on visual contact. The second was to determine whether the type of premises at each address was eligible. To limit the sample to places where crime occurred in public and could reasonably be deterred by police presence, we excluded all residential and most commercial buildings of more than four stories (including two hotels), almost all parking garages and department stores, indoor malls, public schools, office buildings, residential social service institutions (such as homeless shelters), hospitals, police stations, and fire stations. We also excluded parks because almost all were too large to meet the visual contact criterion. Finally, we excluded a few known "magnet phone" locations, at which events occurring elsewhere were routinely reported.

The third goal of the inspection was to determine the visual proximity *between* the cluster and the possible contamination of each site by patrol car presence in the closest neighboring site. The two independent field workers, Michael E. Buerger and Ellen G. Cohn, examined each site and drew what appeared to be logical boundaries. Their separate versions of boundaries for the final hot

---

[3] Some difficulty developed in this process because different definitions of places were used by the City of Minneapolis and by MAPINFO. We were able to reconcile most of these differences, usually by hand-plotting addresses on the computer map, but some 5 percent of the "hot" addresses were left out of our mapping analysis.

spots initially randomized achieved 75 percent agreement. Their reconfigurations followed these general principles:

1. No hot spot is larger than one standard linear street block (although a few exceptions were allowed on the basis of visual sightings on very short blocks).
2. No hot spot extends for more than one half block from either side of an intersection.
3. No hot spot is within one standard linear block of another hot spot (again we made a few exceptions).

The site visits produced a provisional list of 321 maps, with some overlap which we narrowed to a final list of 268 reconfigured clusters (with the ineligible locations excluded).[4] We marked the 268 on a map to make final eliminations based on proximity. Using memoranda about the layout of each site and its proximity to nearby clusters, the principal investigators created a new list of eligible clusters, all of which were required to generate at least 20 hard crime calls in the selection year. This list was also informed by the "soft crime" totals for the selection year (with a minimum of 20), and by an element crucial to the statistical power of the analysis: the percentage change (positive or negative) in the total calls for hard and soft crime from the year ending May 1987 to the year ending May 1988. High variance from year to year could have attenuated the treatment effects, so clusters with greater than 150 percent increases or 75 percent decreases in hard crime calls from one year to the next were excluded from the possible sample. The greatest decrease included in the final sample was 66 percent.

After we made exclusions for variance and the most severe cases of proximity, only 155 hot spots were left. We eliminated four more on the grounds of new data on proximity; one was eliminated because it had become dormant in recent months. At our request, the surviving 150 were randomized by an independent statistician into three treatment groups, which we presented to a planning committee of the Minneapolis Police Department. The committee concluded that the department could not handle 100 target hot spots with adequate dosage to provide a reliable test of the theory, and asked us to reduce the experimental group to 50. The final agreement called for 55 hot spots assigned to extra patrol; thus 110 sites had to be selected for randomization.

We derived the final selection of the 110 sites from the 150 previously identified sites, primarily by taking the top-ranked hot spots in order of volume of hard crime calls. The final 110 were

---

4 Secondary analysts of these data should know that the numbering system for the hot spots in the raw data reflects the surviving members of the provisional list of 365, not the final list of 110.

rerandomized by University of Minnesota statistician Kinley Larntz, despite concerns that about 10 of the clusters would not appear "hot" enough to patrol officers. In the final 110 clusters, the mean number of hard and soft crime calls for service at the active addresses was 182.9 in the selection year, with a minimum of 56 and a maximum of 628.

Characteristics of Hot Spots

The typical hot spot in the final sample of 110 was a group of attached two- and three-story buildings clustered around an epicenter, usually a street corner. Addresses included in the cluster extended in all four directions but only as far as the eye could see from sidewalk corners. These intersections often consisted of a mix of commercial services, usually including food and drink, generally open until late at night. Exceptions to this pattern included low-rise multifamily housing developments and convenience stores. Bus stops and pay telephones were common features of hot spots, as was intensive street lighting.

"Hot" Times

The calls at the 110 spots were concentrated between 7:00 p.m. and 3:00 a.m. We determined this by summing the calls over the selection year by each hour of the day, for both the experimental and the control group. The 7-to-3 window for the experimental group accounted for 51.9 percent of the crime calls; for the control group, this window accounted for 50.5 percent. The 11:00 a.m. to 7:00 p.m. period registered the next highest concentrations, with 32 percent of the experimental group's calls and 33.6 percent of those for the control group. The 3:00 a.m. to 11:00 a.m. period, with the exception of a few sites, registered the fewest calls, with only 16.3 percent of the experimental group's calls and 15.8 percent of the control group's. Thus the experiment was restricted to the period from 11 a.m. to 3 a.m.

Hot Spot Sample Sizes

The sample sizes include several dimensions: the numbers of hot spot clusters, the addresses used to select the clusters, the total number of addresses within those boundaries, numbers of calls for all reasons at those addresses, and the numbers of calls about hard and soft crimes dispatched to those addresses.

The experiment randomly assigned 110 address clusters to treatment and control groups. These clusters contained a total of 677 specific "selection" addresses and intersections (320 experimental group addresses and 357 control), with a mean of six addresses

per site. When all of the addresses included within the boundaries of each hot spot are considered (not only those addresses with three or more calls, as in the selection data cited above), the total was 1,663 (a mean of 15 addresses per spot): 832 addresses in the experimental group and 831 in the control.

During the *baseline* year before the experiment began, these "all-inclusive" clusters produced a total of 19,322 calls for all reasons in the experimental group and 19,693 in the control, or a mean of 355 calls per hot spot. This total constituted 10.8 percent of the 364,365 calls dispatched for all reasons citywide in the one-year baseline period, December 1, 1987 to November 30, 1988. Adjustment for nontraffic calls produced virtually identical proportions.

Treatments

This experiment tests a theory of intensified but intermittent patrol, not a theory of constant, security guard-style presence. The experimental patrol treatment approximates a crackdown-backoff pattern; a police car was not present at the target address clusters at all times. Cars left to answer calls and then returned unexpectedly. They stayed at one spot for as long as an hour or more, or for only a few minutes. Both one-officer and two-officer units were used; foot patrol presence was measured separately. Both officers and observers were given maps of the hot spot boundaries; the addresses generating the most police calls were highlighted in red.

What the officers did while present at the sites varied widely by officer. During an inspection visit at our invitation, George Kelling (1990, personal communication) observed that some were reading newspapers or sunning themselves while sitting on the patrol car, while others were engaging citizens in friendly interaction in community-policing style. The experiment was clearly no test of the content of police presence, only of the amount. To gain police cooperation in achieving the dosage goals, we did not presume to restrict the officers' discretion in *how* to police a hot spot, but only in *how much*.

Random Assignment

The final sample of 110 address clusters was assigned randomly to two groups of 55 by the independent statistician, who used a computerized pseudo-random number generator to allocate the clusters equally to two groups. The allocation was performed in five statistical "blocks," based on natural cutting points within the distribution of hard crime call frequencies. This decision was intended

to increase statistical power by minimizing the differences in variance between the groups. Although blocking results in a loss of degrees of freedom in analysis of experimental effects, it produces a gain over simple randomization by maximizing the equivalence of the groups. Further, a comparison of randomization by pairs with randomization in five blocks showed little difference in statistical power.

Dosage

After extensive debate, the police department committed itself to (but never fully achieved) a goal of three hours a day of patrol presence at each of the 55 target hot spots. The dosage, based on the above analysis of "hot times," was to be divided evenly between the 11-7 and the 7-3 time periods, and was to be provided seven days a week. To enhance the power of the experiment (Weisburd 1993), our goal (which was largely achieved) was to keep dosage levels as consistent as possible. We encouraged this by giving patrol managers weekly reports on the dosage levels reported by officers in their official logs. These reports were supplemented by a monthly report on the amount of dosage recorded by our field observers. When some spots appeared in the logs to be receiving more dosage than others, we asked patrol supervisors to assign less time at those spots and to order more time at the locations receiving less logged dosage.

The independent observations by our field staff of 16 observers and three supervisors were limited to the 100 most active control and experimental spots; the five "coolest" spots in each treatment group were eliminated from the observations to maximize measurement of the places producing the largest volume of crime. The observations covered a total of 75 hours per hot spot over the course of the year. All observations were made between 7:00 p.m. and 3:00 a.m. The 7,542-hour sample thus constituted 2.6 percent of all hours on that period over 365 days times 100 locations (292,000 hours).

The observation sample was divided equally into 13 periods of 28 days each for each hot spot. Observations were conducted in a total of 6,465 blocks of 70 minutes each. Each of the 13 28-day periods contained 497 observation blocks, or about five per hot spot. A total of 3,232 observation blocks were conducted for the 50 experimental hot spots, and 3,233 for the 50 observed controls.

Observers were trained to use a systematic observation instrument that employed separate sections for observations of uniformed officers and of crime and disorder. Both sections were structured chronologically so that each entry had a start and a finish time, as

did the entire observation period. Entries registering official presence included "drive-throughs" and longer stays of police in cars and on foot, private security guards, fire truck and ambulance personnel, and whether and how long police left their cars or entered buildings. Observations of crime and disorder included an array both criminal offenses and offenses against conventional civility, as noted below.

Outcome Measures

We collected two primary outcome measures: calls about crime and observed disorders. The hot spots were selected on the basis of telephone calls about criminal activity reported by the public—as distinct from dispatchers' records of events reported by police officers over the radio, which also can generate a "call" record. Therefore citizen calls should be treated as the primary outcome measure. Calls about "soft" and "hard" crime were counted for the full 24-hour day, not only the 16 hours in which the experiment was operational, for two reasons. One was theoretical, based on our conception of general deterrence as including "residual" effects even when police patrols are not present (Sherman 1990). The other reason was statistical: we included the full 24 hours in order to increase the power of the test by using higher base rates of crime calls in each hot spot.

The other outcome measure was a more direct measure of crime than citizen calls, although it necessarily lacked baseline data for sample selection. Systematic observation data on crime and disorder in the evening observation hours coded each incident of fights, drug sales, apparent solicitation for prostitution, playing of loud music or shouting, rummaging through garbage cans, urinating, and other offensive "signs of crime" (Skogan 1990). The data even included two minor assaults on an observer sitting in a parked car.[5]

We planned to analyze the police call data by comparing Time 1/Time 2 differences between the two groups, and to analyze the observations at Time 2 only. Time 1 is the 12 months preceding the experiment; Time 2 is the 12 months of the experiment (December 1, 1988 through November 31, 1989).

---

5 This issue is important for systematic observation of hot spots. Observers were instructed to always observe hot spots from inside an auto, and to leave if they ever felt there was any question about their safety. Both assaults were committed through an open window while the (same) observer was smoking a cigarette. Aside from these assaults, the risk to the observers appeared quite low, but this situation could be different in a city with "hotter" hot spots.

Analysis of Statistical Power

The statistical power of a test "is the probability that it will lead to a rejection of the null hypothesis" (Cohen 1977:4), or the odds of detecting a statistically significant result in an experiment (at each significance level) given a true difference between experimental and control groups. We computed the power of our incidence measure using the selection-year data for the final 110 hot spots with a one-tailed 10 percent test. We used a one-tailed test because of our strong hypothesis that patrol presence reduces crime. We chose a 10 percent significance level because police executives are more interested in size of an effect than in the exact odds that the effect is due to chance. On the basis of tables provided by Cohen (1977), and assuming a standard deviation of 33.5 percent for total crime and a 10 percent significance level, we estimated that we had an 85 percent chance of gaining a significant finding in our experiment if the true impact of the treatment was about 15 percent. This level of power exceeds the .80 threshold suggested by Cohen (1977) for powerful experimental designs.

Summary of the Design

We designed this experiment to test the hypothesis that substantial increases in police patrol in high-crime hot spots could reduce crime reported and observed in those spots. We selected the hot spots on the basis of calls for service and visual proximity. The independent variable was assigned at random to half of a group of 110 hot spots constituting a universe of all address clusters meeting certain minimal levels of "hard" and "soft" crimes, as well as stability over two years in calls for police service for those types of incidents. We measured the independent variable by police logs and by independent observation of the 50 most active hot spots in each group of 55. The dependent variable was measured by police calls for service and by independently observed incidents of crime and disorder.

For 6 1/2 months, the design was implemented as planned. What happened then to modify the design produced results generally consistent with the hypothesis, even while it reduced the intended statistical power by cutting the anticipated experimental period almost in half.

RESULTS

Independent Variable: Observed Differences in Dosage

From December 1, 1988 to November 30, 1989, the observers counted 34,416 police unit-minutes in the 50 observed experimental

hot spots and 14,765 unit-minutes in the 50 observed control hot spots, a pooled ratio of 2.3 to 1. The difference in mean police presence per hot spot was slightly lower at 1.99 to 1, with $\bar{X} = .149$ police unit-minutes per minute of observation in the 50 observed experimental hot spots and $\bar{X} = .0748$ police-unit minutes per minute of observation in the 50 control hot spots. A "unit-minute" refers to the number of minutes each police unit spent in each location; "units" include one-officer marked cars, two-officer marked cars, and one- or two-officer foot patrols. Whenever a police unit entered the boundaries of the hot spot, the observer started the clock counting for the minutes of that unit's presence. The count ended when either the unit or the observer left the hot spot. The minutes present for each unit sometimes overlapped, so that unit-minutes divided by observation minutes cannot be taken as a prevalence measure of any police presence at all.

Compliance with the experimental protocol can be estimated by analyzing the ratio of unit-minutes to all observed minutes in each of the 100 observed hot spots. Using a criterion of one unit-minute of observed police presence for every 10 minutes of observations as the threshold for defining an "experimental" case, we find five hot spots assigned to the experimental group which failed to receive that level of dosage and four hot spots assigned to the control group which did receive that amount. Thus the "misassignment" or "crossover" rate in traditional experimental terms is 9 percent, or 9 out of the 100 observed cases. This rate is moderate for randomized trials generally, and better than the rate in most police experiments (see Dennis 1988; Weinstein and Levin 1989). Otherwise the hot spots received highly similar within-group dosage levels: 46 of the 50 experimental hot spots received 1.3 to 1.7 minutes of patrol per 10 minutes of observation, and 40 of the controls received either .7 or .8 police minutes per 10 observed minutes.

The summer design breakdown.

Although the mean unit-minutes across hot spots within treatment groups were relatively homogeneous, the pooled ratio between experimental and control unit-minutes varied widely by calendar month. The ratio began at 2.6 to 1 in December and fell in January to 2 to 1, where it remained until March. At that time it rose to 6 to 1 and then fell to about 2.5 to 1 in April through June. The ratio then plummeted to 1.2 to 1 in August, and rose in September to a plateau of 2.8 to 1, and remained at that level for the rest of the experiment (see Figure 1). The police logs reflect the same pattern, declining from an average of just under three hours per day in the experimental hot spots from February through May to only two

hours in July and August, and rising again in the autumn. Although the observed police unit-minutes ratio exceeded 2 to 1 for every month except August, the disruption of the experiment during the summer peak in call load (and vacation time) for police complicates the interpretation of any differences in outcomes over the entire one-year period, leaving only 6.5 months of a fully implemented design.

Figure 1. Ratio of Experimental to Control Minutes of Observed Police Presence, by Month

Ratio of Minutes of Police Presence
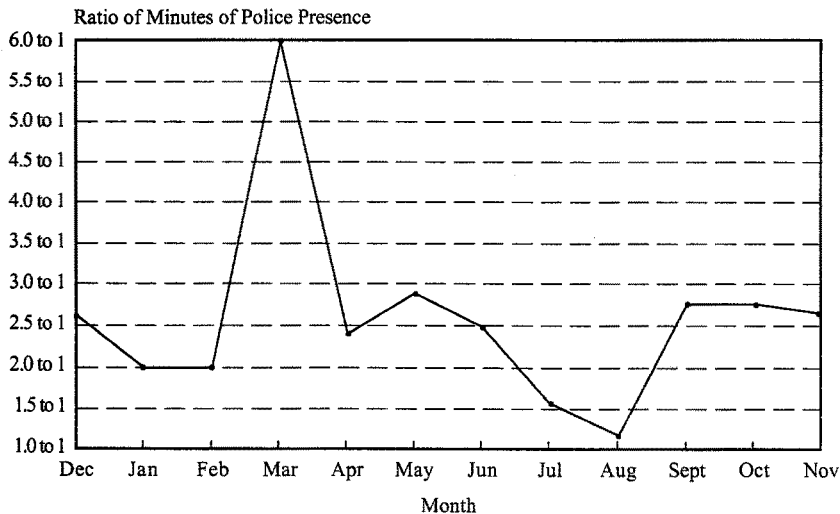
Dependent Variable 1: Differences in Calls about Crime

The virtual disappearance of a difference in patrol dosage between experimental and control groups in the summer months raises several options for analysis. These options are further complicated by an outcome measurement problem caused by the introduction of a new computer-aided dispatch (CAD) system from October through November 10, 1989. During that period, errors and missing data made the calls about crime an unreliable indicator. One option—perhaps the simplest—is to analyze the period from December 1 through June 15, when the police logs show the most consistent and most uninterrupted implementation of the experiment throughout the 16-hour target zone. Another option is to cut off analysis at July 31, before the only month in which observational data show virtually no difference in dosage (a period in which the overall ratio is 2.5 to 1). A third option is to analyze the full

year, despite the six weeks of CAD measurement problems in October and November. A fourth option is to analyze the full year minus the period of suspect CAD data.

We find the July 31 cutoff to be the most appropriate test of the hypothesis because that date is the last date on which the experiment was minimally implemented as planned. Because others may disagree, however, we present the data for all four of the time periods defined above.[6]

Table 1 presents the raw data for differences in hard, soft, and total citizen calls about crime for each of the four periods, as well as the significance levels for the mean Time 1 to Time 2 differences per hot spot between treatment and control groups as calculated from a mixed model ANOVA test taking randomization block into account. It shows that total crime calls and calls about soft crime increased from the baseline to the experimental year in both treatment and control groups, while calls about hard crime decreased in both groups from the baseline to the experimental year. Thus the analysis centers on the *differences of differences* between the baseline and the experimental years, comparing experimental with control hot spots.

---

6 We used a mixed-model analysis of variance, taking into account the effects of randomization block and treatment group as well as the interaction between block and treatment group. Each significant finding was subjected to tests for stability. We examined the effects of removing and including blocks of cases, of transforming the distributions of events, and of results obtained by using less powerful rank-ordering techniques, including the nonparametric combined independent Mann-Whitney rank order test. All tests produced the same results in the call analysis.

## Table 1. Crime Calls by Time Period and Treatment Group

Time Period	Hard Crime		Soft Crime		Total Crime	
	Experimental	Control	Experimental	Control	Experimental	Control
June 15						
Baseline year	1,469	1,394	3,544	3,590	5,013	4,984
Experimental year	1,377	1,374	3,919	4,542	5,296	5,916
Absolute change	−92	−20	375	952	283	932
1-Tailed P Value	.27		.047*		.054*	
July 31						
Baseline year	1,893	1,798	4,638	4,693	6,531	6,491
Experimental year	1,776	1,793	5,155	5,909	6,931	7,702
Absolute change	−117	−5	517	1,216	400	1,211
1-Tailed P Value	.20		.046*		.049*	
November 30[a]						
Baseline year	2,533	2,432	6,523	6,644	9,056	9,076
Experimental year	2,455	2,419	7,116	8,049	9,571	10,468
Absolute change	−78	−13	593	1,405	515	1,392
1-Tailed P Value	.33		.046*		.058*	
November 30[b]						
Baseline year	2,873	2,741	7,396	7,664	10,269	10,405
Experimental year	2,754	2,700	8,163	9,016	10,917	11,716
Absolute change	−119	−41	767	1,352	648	1,311
1-Tailed P Value	.31		.155		.159	

* $p < .10$
[a] Excludes period from 10/1 to 11/10.
[b] Includes period from 10/1 to 11/10/89.

Figure 2 shows that the predicted effect, on total crime calls, of the reduced difference in patrol dosage appears in August, on schedule. At that time the experimental group fails for the first time to show a more favorable absolute difference, in calls from the same period in the prior year, than the control group. In every month before August, when the experimental group received far more police presence than the control group, the Time-1-to-Time-2 change in total calls had been more favorable for the experimental group. The August violation of that pattern disappeared in September but returned in October, when the data on calls became questionable. The violation of the predicted difference disappeared again in November, when the new CAD system was thought to be reliably established.

Table 2 presents the absolute baseline-experimental percentage differences and the difference of those differences between experimental and control hot spots as computed by a mixed-model analysis of variance using the five-block design. The effect of increasing patrol is greater on total and soft crime calls than on hard crime. Soft crime effects are strong in every period except the full year including the CAD changeover errors; they range in magnitude of relative percentage differences (experimental group baseline to experimental year percentage change minus control group baseline to experimental year percentage change) from 7 percent for the

**Figure 2.   Absolute Differences From Baseline to
Experimental Year in Total Crime Calls by
Month and Treatment Group**
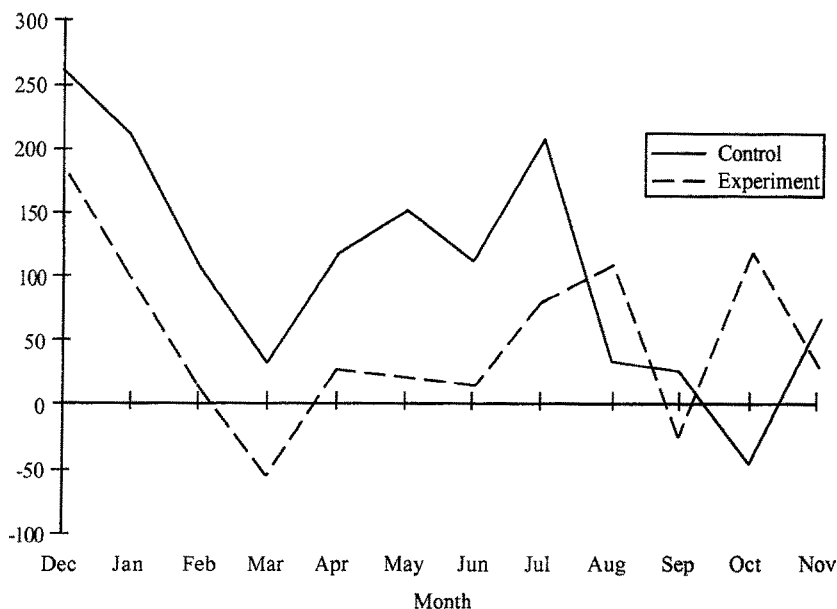
full year to 16 percent for the period ending June 15. The effects for total crime calls are similar but attenuated because the soft crime calls account for most of the total crime calls.[7]

The concept of percentage difference is presented conservatively; we compare absolute percentage changes rather than the percentage difference of percentage differences. That is, even for the full year we could say that the increase in soft crime calls was 75 percent greater in the control group than in the experimental

_____

[7] Whether these differences are a function of a displacement of citizens' calls onto the officers already present at hot spots is an interesting question; it reveals the failure of this design to eliminate the problem of interpreting the effects of police presence on citizens' propensity to call police, given a reason to do so. Adding in the police-generated calls made at the hot spot addresses is one proposed solution, for which we thank Professor Carl Klockars. We find this solution unsatisfactory, however, because it cannot distinguish events that citizens report to police at the scene (and would have called 911 to report, had no police been there) from events that police call in about (such as car checks), and about which citizens would never have called 911. Because of the small number of minutes when police are present, even in the experimental hot spots, any displacement of citizens' calls seems likely to be minimal, whereas the generation of police-initiated calls while they are assigned to the hot spots, as we know from direct observation in the police cars, is quite substantial. As predicted by both interpretations of this indicator, the addition of police-generated crime calls to the citizen-generated calls creates no significant differences between treatment groups in any of the time periods or crime types (data not displayed). The addition of an hour per day of patrol presence accounts at most for one-eighth of the 50 percent of calls generated between 7 p.m. and 3 a.m. or 6.25 percent of all calls. Thus the maximum displacement of citizens' calls from 911 to police would seem to be less than half of the measured crime reduction of 13 percent or more relative to the control trend.

**Table 2. Percentage Changes of Crime Calls from
Baseline to Experimental Year, by Time Period,
Treatment Group, and Significance Levels of
Mixed-Model ANOVA Tests**

Time Period	Hard Crime			Soft Crime			Total Crime		
	Exp.	Control	Difference	Exp.	Control	Difference	Exp.	Control	Difference
June 15									
Percent change	−6.3	−1.4	−4.9	10.6	26.5	−15.9	5.6	18.7	−13.1
July 31									
Percent change	−6.2	−.3	−5.9	11.1	25.9	−14.8	6.1	18.7	−12.6
November 30[a]									
Percent change	−3.1	−.5	−2.6	9.1	21.1	−12.0	5.7	15.3	−9.6
November 30[b]									
Percent change	−4.1	−1.5	−2.6	10.4	17.6	−7.2	6.3	12.6	−6.3

[a] Excludes period from 10/1 to 11/10.
[b] Includes period from 10/1 to 11/10.

group (17.6 percent divided by 10.4 percent). By subtracting the percentage differences rather than dividing them, we focus the analysis on the magnitude of crime differences associated with more patrol rather than on its proportionate effect.
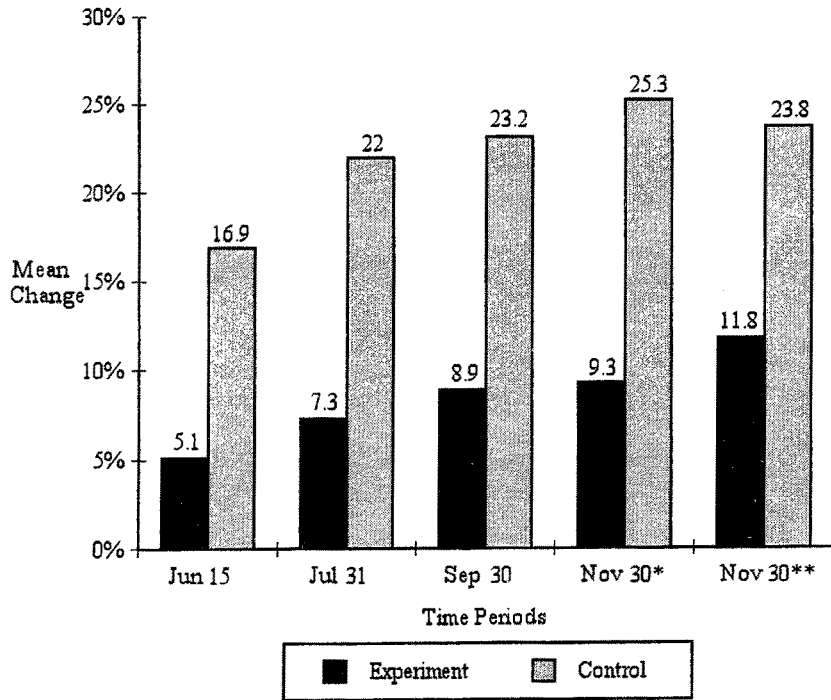
Figure 3 reports and illustrates the mean Time 1/Time 2 differences in calls for the experimental and the control groups, using different cutoff dates for the experiment. It is clear that no matter what cutoff date is selected, the increase in citizen calls in the 55 control hot spots is substantially greater than in the 55 experimental hot spots. The absolute size of the difference at any one hot spot is quite modest, however—about one fewer crime call per month.

*Dependent Variable 2: Differences in Observed Crime and
Disorder*

The disorder analysis shows the most striking differences between the experimental and the control groups of any analyses. Table 3 displays the percentage of minutes of observations in different time periods in which disorderly public conduct was observed, by treatment group.[8] For the entire experimental period, we find a significant relative difference of 25 percent less disorder in the experimental than in the control group. For the two periods in which the experiment had the greatest integrity (ending June 15 or July 31), the effect was even stronger: half as much disorder was observed in the experimental group as in the control. The absolute difference of only 2 percent of all observed minutes versus 4 percent

---

[8] Because more than one disorder could have occurred simultaneously, Table 3 actually represents a ratio between observed minutes of all disorders and minutes of observations.

Figure 3.   Percentage Change From Baseline to Experimental Year in Total Crime Calls Per Hot Spot by Treatment Group and Period.

Chart:

- Y-axis: Mean Change, 0% to 30%
- X-axis: Time Periods — Jun 15, Jul 31, Sep 30, Nov 30*, Nov 30**
- Experiment (black): 5.1, 7.3, 8.9, 9.3, 11.8
- Control (gray): 16.9, 22, 23.2, 25.3, 23.8

Legend: ■ Experiment ▨ Control

* - Excludes period from 10/01/89 - 11/10/89

** - Includes period from 10/01/89 - 11/10/89

reflects a difference, in odds of encountering a disorder, between 1 in 50 and 1 and 25. For a resident or user of any cluster of addresses, this difference is noticeable and substantial.

This large relative difference is not due simply to a deterrent effect on disorder while police are present. Only 6 percent (209 of 3,513) of observed disorder events began while police were present across the entire observed sample. Koper (1995) reports significant differences in observed disorders between experimental and control groups when police are *not* present—up to 65 percent less criminal disorder in the experimentals.

An analysis of 13 specific types of disorder for the entire year shows that the greatest effects (in which ratios of control disorder incidents to experimental disorder incidents exceeded 1.5 to 1) were on the categories of person down (on the ground), drug activity, vandalism, solicitation for prostitution, and assault. We found no difference, however, in observations of persons apparently drunk or

Table 3. Minutes of Disorder Observed in Experimental and Control Groups Compared with ANOVA Tests Controlled for Blocking

Period and Group	Minutes of Disorder	Minutes of Observations	Mean Ratio Per Hot Spot
Entire Year			
Experimental	5,855	225,991	.026
Control	8,623	226,295	.038
1-Tailed P Value			.022*
Until 6/15/89			
Experimental	2,267	121,363	.019
Control	4,493	122,736	.037
1-Tailed P Value			.006*
Until 7/31/89			
Experimental	3,545	148,617	.024
Control	5,915	149,889	.040
1-Tailed P Value			.007*

* p < .10

drugged, the largest single category of disorder (but perhaps the one theoretically least deterrable by police presence).

Table 3 displays the difference between experimental and control groups in observed disorder ratios. One-tailed P values are derived from ANOVA tests taking into account the five blocks used for the original random assignment of all 110 hot spots, only 100 of which were observed. All ten unobserved spots (five experimentals and five controls) were in the same randomization block because the blocks were stratified by volume of hard crime calls. That block is fortunately the largest, with 58 hot spots, of which observations on ten (17%) are missing. The analysis simply treats those cases as missing data. No matter what time period we examine, these experimental year treatment group differences in observed disorder ratios are highly unlikely to be due to chance sampling fluctuations.[9]

CONCLUSIONS

These results show clear, if modest, general deterrent effects of substantial increases in police presence in crime hot spots. Just as police strikes reveal major increases in crime due to major reductions in police presence (Makinen and Takala 1980; Russell 1975), our findings show that the difference in crime is proportionate to the difference in police. If urban police agencies decided to assign

---

[9] The P values in Table 3 are derived from an ANOVA design in which the effects of treatment group and block are included. The interaction between treatment type and block is not statistically significant, and is excluded from the model.

even higher priority to hot spot patrols, the magnitude of the crime reductions might be even greater.

This conclusion, however, presents two problems. One is that the effects of police on crime in hot spots may be attenuated by displacement of that crime to other locations. Absent any test of that interpretation, we cannot rule out the claim that more police will push crime around rather than preventing it. Yet in light of the strong conclusions drawn about the Kansas City Preventive Patrol Experiment (Kelling, et al 1974), even these results falsify the claim that patrol has no effect on crime at all.

Although we cannot conclude that these findings show a general deterrent effect of police presence throughout the community, we can claim evidence of place-specific "micro-deterrence." Even if police patrol pushes the crime elsewhere, it has been generally deterred by police presence in that location. The concept of deterrence is based on a rational calculation of risks and benefits. The prevention of crime and disorder in experimental hot spots, even when police are not there, is consistent with the hypotheses of apprehension and punishment in that place. This may be the same mechanism that causes displacement to a location where the fear of punishment is less, but it also fits the micro-general deterrence model precisely.

A second, different problem in recommending more hot spot patrols is that police may find directed patrol distasteful. The deterrent findings suggest that the more the time police stay in a hot spot, the less opportunity they will have to exercise police powers. This is good for the community but can be boring for the police. Rather than preventing crime by keeping hot spots cool, most police would prefer to catch criminals after crime has already occurred and the harm has been done. Prevention lacks glamour; apprehensions offer the excitement of the chase. A substantial change to a community policing philosophy could make hot spot patrols more interesting, especially if police leave their cars and talk to frequent users of the hot spots. But historically the resistance to such a change has been formidable.

More detailed analysis suggests how to minimize police resistance without a major philosophical change. The greatest deterrent effect may be produced not by police staying in the same hot spot for extended periods, but by police roving from hot spot to hot spot, staying in each for only a limited time. In this issue, Koper (1995) reports a curvilinear effect of the duration of police presence in hot spots on the amount of time that elapses until the first disorder or crime event is observed after police leave. The optimal length of a hot spot patrol appears to be about 12 minutes. This should be well

within the police boredom threshold, allowing them to move on to the next hot spot to see who might be causing trouble upon their arrival.

This experiment remains unreplicated, and may be limited in external validity to the time and place where it was conducted. We urge caution in generalizing its results to other settings. At the same time, we conclude that the experiment offers a more powerful and more externally valid test of the patrol deterrence hypothesis than the Kansas City experiment. At the very least, it is time for criminologists to stop saying "there is no evidence" that police patrol can affect crime.

REFERENCES

Barr, R. and K. Pease (1990) "Crime Placement, Displacement, and Deflection." In M. Tonry and N. Morris (eds.) *Crime and Justice*, Vol. 12, pp.227-318. Chicago: University of Chicago Press.

Boydstun, J. (1975) *San Diego Field Interrogation: Final Report*. Washington, DC: Police Foundation.

Buerger, M.E., E.G. Cohn, and A.J. Petrosino (1995). "Defining the Hot Spots of Crime: Operationalizing Theoretical Concepts for Field Research." In D.A. Weisburd and J.E. Eck (eds.), *Crime and Place: Crime Prevention Studies*, Vol. 4. Forthcoming. Monsey, NY: Criminal Justice Press.

Chaiken, J. (1978). "What is Known About Deterrent Effects of Police Activities". pp.109-36. In J. Cromer, ed., *Preventing Crime*. Beverly Hills, California: sage.

Chaiken, J., M.W. Lawless, and K.A. Stevenson (1974). "The Impact of Police Activity on Subway Crime." *Urban Analysis* 3:173-205.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. NY: Academic Press.

Cornish, D.B. and R.V. Clarke, eds. (1986). *The Reasoning Criminal: Rational Choice Perspectives on Offending*. New York: Springer-Verlag.

——— (1987). "Understanding Crime Displacement: An Application of Rational Choice Theory." *Criminology* 25:933-47.

Dennis, M.L. (1988). "Implementing Randomized Field Experiments: An Analysis of Criminal and Civil Justice Research." Doctoral dissertation, Northwestern University, Department of Psychology.

Farrington, D.P. (1982) "Randomized Experiments on Crime and Justice." In M. Tonry and N. Morris (eds.) *Crime and Justice: An Annual Review of Research*, Vol.4, pp.257-308. Chicago: University of Chicago Press.

Felson, M. (1994) *Crime and Everyday Life*. Thousand Oaks, CA: Pine Forge Press.

Fielding, H. ([1751] 1977) *An Enquiry into the Causes of the Late Increase of Robbers*. Montclair, NJ: Patterson-Smith.

Fienberg, S., K. Larntz, and A.J. Reiss Jr. (1976) "Redesigning the Kansas City Preventive Patrol Experiment" *Evaluation* 3:124-31

Freiman, J.A., T.C. Chalmers, H. Smith Jr., and R.R. Kuebler (1978) "The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Controlled Trial: A survey of 71 'Negative' Trials." *New England Journal of Medicine* 299:690-4.

Gottfredson, M. and T. Hirschi (1990) *A General Theory of Crime*. Stanford: Stanford University Press.

Green, L. (1995) "Cleaning Up Drug Hot Spots in Oakland, California: The Displacement and Diffusion Effects." *Justice Quarterly*, This issue.

Kelling, G., A.M. Pate, D. Dieckman, and C. Brown (1974a) *The Kansas City Preventive Patrol Experiment: Technical Report*. Washington, DC: Police Foundation.

——— (1974b) *The Kansas City Preventive Patrol Experiment: Technical Report*. Washington, DC: Police Foundation.

Klockars, C., ed. (1983) *Thinking about Police*. New York: McGraw-Hill.

Koper, C. (1995) "Just Enough Police Presence: Reducing Crime and Disorderly Behavior by Optimizing Patrol Time in Crime Hot Spots." *Justice Quarterly*, this issue.

Larson, R.C. (1975) "What Happened to Patrol Operations in Kansas City?" *Journal of Criminal Justice* 3:267-97.

Makinen, T. and H. Takala (1980) "The 1976 Police Strike in Finland." *Scandinavian Studies in Criminology* 7:87-106.

Matthews, R. (1986) *Policing Prostitution: A Multi-Agency Approach*. London: Middlesex Polytechnic Centre for Criminology.

McNamara, J. (1974) "Foreword" In Kelling, et al, 1974a.

Pate, A.M. and G.L. Kelling (1975). "A Response to 'What Happened to Patrol Operations in Kansas City?'" *Journal of Criminal Justice* 3:299-330.

Press, S.J. (1971) *Some Effects of an Increase in Police Manpower in the 20th Precinct of New York*. New York: New York City RAND Institute.

Reiss, A.J., Jr. (1985) *Policing a City's Central District: The Oakland Story*. Washington, DC: National Institute of Justice.

Russell, F. (1975) *A City in Terror: 1919—The Boston Police Strike*. New York: Viking.

Schnelle, J.F., R.E. Kirchner Jr., J.D. Casey, P.H. Uselton Jr., and M.P. McNees (1977) "Patrol Evaluation Research: A Multiple-Baseline Analysis of Saturation Police Patrolling during Day and Night Hours." *Journal of Applied Behavior Analysis* 10:33-40.

Sherman, L.W. and D. Weisburd (1995). "Does Patrol Prevent Crime? The Minneapolis Hot Spots Experiment" In Koicki Miyazawa and Setsuo Miyazawa (eds.), *Crime Presentation In The Urban Community*. Boston: Kluwer.

Sherman, L.W. (1986) "Policing Communities: What Works?" In A.J. Reiss Jr. and M. Tonry (eds.), *Communities and Crime*, pp.343-86. Chicago: University of Chicago Press.

——— (1990) "Police Crackdowns: Initial and Residual Deterrence." In M. Tonry and N. Morris (eds.), *Crime and Justice*, Vol. 12, pp, 1-48. Chicago: University of Chicago Press.

Sherman, L.W., P.R. Gartin, and M.E. Buerger (1989) "Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place." *Criminology* 27:27-55

Skogan, W. (1990) *Disorder and Decline*. New York: Free Press.

Skolnick, J. and D. Bayley (1986) *The New Blue Line*. New York: Free Press.

Weinstein, G.S. and B. Levin (1989) "Effect of Crossover on the Statistical Power of Randomized Studies." *Annals of Thoracic Surgery* 48:490-5.

Weisburd, D.A., with Anthony Petrosino and Gail Mason (1993) "Design Sensitivity in Criminal Justice Experiments: Reassessing the Relationship between Sample Size and Statistical Power." In M. Tonry and N. Morris (eds.), *Crime and Justice*, Vol. 17, pp.337-89. Chicago: University of Chicago Press.

Weisburd, D.A. and L. Green (1995) " Measuring Immediate Spatial Displacement: Methodological Issues and Problems." In D.Weisburd and J.E. Eck (eds.), *Crime and Place: Crime Prevention Studies*, Vol.4, Forthcoming Monsey, NY: Criminal Justice Press.

Zimring, F. (1978) "Policy Experiments in General Deterrence, 1970-75." In A. Blumstein, J. Cohen, and D. Nagin (eds.), *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*, pp.140-73. Washington, DC: National Academy of Sciences.